| Bonn. zool. Beitr. | Bd. 50 | H. 4 | S. 305–311 | Bonn, Januar 2003 |
|---|---|---|---|---|

# Is not Hennig's method of producing cladograms as defensible as those derived from parsimony algorithms?

R. H. L. Disney

Abstract. In a recent paper Härlin (1999) argued that the evolutionary tree for a taxon that phylogenetic systematics seeks to discover has logical priority over taxonomic characters, and that it follows that congruence among character states therefore has priority over all other considerations. This opinion is challenged and the approach advocated by Hennig (1966) commended as being at least as useful as a means of generating interesting hypotheses.

Key words. Phylogeny, cladistics, systematics, congruence, parsimony.

## Introduction

The two dominant preoccupations of taxonomists are alpha taxonomy, concerned with the recognition and description of species, and beta taxonomy, concerned with arranging species into a hierarchical scheme of classification. It is currently estimated that a large majority of the species on planet earth remains unknown to science. However, not only is the funding for alpha taxonomy in decline but it is now easier to get a paper discussing theoretical aspects of beta taxonomy published than one advancing knowledge of alpha taxonomy (Disney 1999). This is despite the fact that most of these theoretical considerations have been discussed many times before, as indicated in standard texts (e.g. Mayr & Ashlock 1991).

With regard to beta taxonomy, practising taxonomists, as opposed to some theoreticians, generally recognize that the designation of a genus embodies a set of opinions about the affinities of species and the status of the clusters of species we have constructed to reflect those opinions. The process of clustering related species, systematization, is in principle a process of scientific discovery through the erection of testable hypotheses. The assignment of a rank (genus, subgenus or whatever) to a cluster of species, categorization, owes as much to historical accident and prejudice as to scientific insight. A classification, therefore, is partly a reflection of the results of scientific discovery and partly the result of arbitrary decisions. The latter introduce an element of historical constraint, which needs to be respected in the interests of nomenclatural stability. It is too often forgotten that a classification is meant to be a system of reference to aid communication. Continual changes in names, or the concepts signified by them, impair effective communication. Useful theoretical debates are therefore best restricted to the methods used in systematization. Since Hennig's (1966) classic text, most taxonomists have come to regard the recognition of synapomorphies (shared derived character states) as the key to systematization. The continuing debates that concern practising taxonomists are therefore those concerned with the procedures employed in the recognition of synapomorphies.

## Recognition of Synapomorphies

Hennig (1966) insisted that the systematization of taxa should be based on synapomorphies. He thus rejected phenetics, which I have defined elsewhere (Disney 1983) as "the clustering of species by statistical procedures which treat a limited selection of characters as being of equal weight. The resulting classification is a function of the characters selected. A different selection is likely to produce a different classification. Network classifications are more likely to emerge from phenetic analysis than are cladistic classifications". I would merely add that if one employs some system of weighting certain character states (in order to produce a cladogram rather than a network) then one needs to justify one's criterion for selecting those character states preferentially.

The recognition of homologies and the resolution of the polarities of postulated transformation series have remained the key to the recognition of apomorphic character states. The term homology evidently means different things to different authors. I have previously (Disney 1983) defined it thus – "a character state is a homology of another character state when they both form part of the same evolutionary transformation sequence. They will always have the same basic structure. They may not have the same function and do not necessarily resemble each other". This is essentially in accord with a definition such as that of Wägele (1996) – "an evolutionary novelty is an inherited change from a previously existing character. The novelty is the homologue of the previously existing character in an ancestor/descendant relationship". In the case of morphological features, Hennig (1966) considered that a few well founded postulated synapomorphies were likely to prove a more reliable basis for constructing a cladogram than a plethora of character states whose status as postulated synapomorphies was not supported by explicit evidence and argument. However, a growing number of taxonomists have abandoned Hennig's viewpoint and advocate instead that we should determine synapomorphies retrospectively through the application of parsimony algorithms to large databases of character states tabulated against taxa (e.g. Scotland 1992, Yeates & Wiegmann 1999). In effect congruence becomes the criterion for the preferential selection of character states. The states thus selected are then inferred to be the synapomorphies being sought. Indeed some authors, such as Härlin (1999), explicitly take this to its logical conclusion by arguing for the priority of character congruence over all other considerations. Such a procedure amounts to a modified form of phenetics, despite the latter now being recognized as an essentially arbitrary method of clustering species. Such authors bypass, rather than address, the overwhelming evidence (e.g. Moore & Willmer 1997) for the prevalence of convergent evolution. In seeking the preferred cladogram for a data matrix based on morphological features, they assume that the only reliable way to recognise homoplasy is through finding the most parsimonious solution and that those states not supporting this cladogram must, therefore, be the homoplasies (= the 'analogies' of authors such as Wägele 1996). However, the greater degree of linkage between the character states employed the more will procedures based primarily on congruence run the risk of amounting to circular arguments. This is because pleiotropy and polygenic control of the character states under consideration will be the norm for the majority (if not all) of these states.

## Molecular data

In the case of DNA sequence data it has usually been assumed that the degree of linkage will tend to be reduced to a level where reliance on congruence can be employed without thereby inevitably becoming an exercise in circular argument. Indeed it is difficult to avoid parsimony algorithms or other statistical procedures (such as maximum likelihood and neighbour-joining algorithms for sorting taxa in relation to tabulated character states) with the analysis of DNA sequence data because of the dearth of alternative procedures. Perhaps, it also needs emphasizing that such sequence-based cladograms portray affinity in terms of genetic similarity rather than in terms of the recency of common origin that is portrayed by a cladogram based on morphological data and using Hennig's classic approach to constructing the cladogram. While it is reasonable to assume that shared sequences represent synapomorhy, there always remains a possibility of such sequences being homoplasies. However, as we learn more about the functions of these sequences we should be able to detect such cases with greater confidence. While these two sorts of affinity are likely to be very closely correlated, they will also differ to a greater or lesser extent. Prompted by Härlin's recent paper in particular these issues are briefly discussed further below.

## Discussion

Härlin (1999) argues that the evolutionary tree for a taxon that phylogenetic systematics seeks to discover has logical priority over taxonomic characters, and that it follows that congruence among character states therefore has priority over all other considerations. His somewhat theoretical arguments in favour of this seem to have little relation to taxonomic practice.

Of course the true evolutionary tree for a taxon is awaiting discovery and therefore, in this trivial sense, has logical priority in that it is the aspiration to discover this tree that motivates phylogenetic systematics in the first place. In practice, however, the discovery of the tree is the goal of an investigation rather than the starting point.

In practice the starting point of all taxonomic endeavour is the examination of individual specimens of organisms. These can all be demonstrated to be unique. It follows that any clustering of these individual organisms into taxa necessitates focusing on some character states while ignoring others. At the level of alpha taxonomy we cluster individuals into putative species (which I will assume, for the sake of argument, are sexually reproducing species) on the basis of biological data (e.g. pairs caught mating) and selected morphological character states. In the process we relegate some character states to the status of being examples of individual, infraspecific, variation. Other differences we may designate as being cases of sexual dimorphism. As we proceed to beta taxonomy we continue to select some characters and to ignore others as we cluster the species into a hierarchy of taxa.

The key issue for beta taxonomy has always been the justification of the criterion that leads us to select some character states at the expense of others. Ever since Aristotle taxonomists have recognized that some characters could be weighted preferentially as being indicative of something significant while others only emerged as being significant because they were found to correlate with many other character states. Thus Aristotle weighted red blood, as he considered this to be indicative of the 'essence' of the taxon under consideration. His approach culminated in the achievements of typological taxonomy, which, however, failed to justify its various notions of the 'essence'. The Darwinian revolution replaced the notion of the 'essence' with that of evolutionary affinity. Evolutionary taxonomists then proceeded to try to identify character states indicative of such affinity. However, many rival claims emerged to the extent that some workers shifted attention back to Aristotle's other criterion – the correlation of character states (or congruence in Härlin's vocabulary). Modern phenetics gave us many useful quantitative, statistical, procedures for clustering taxa. However, such methods are based on at least two assumptions that are difficult to defend (see below) and furthermore the results are always at risk of distortions due to the characteristics of the algorithms employed rather than the realities of nature.

While it was the choice of criteria for the preferential selection of character states that was the subject of debate prior to the general acceptance of synapomorphy, now the debate has shifted to the methods we employ to recognise synapomorphies. This shift gave rise to a revival of procedures pioneered by phenetics and their adaptation as parsimony algorithms employed as a means of recognising homoplasy and synapomorphy. For certain types of data such algorithms (e.g. those included in computer programmes such as PAUP) may be the only option available, despite their limitations.

The most questionable assumption of parsimony/congruence algorithms is that each piece of information (each datum) is equivalent in weight to every other datum in one's database of character states or a selected subset of it (as in the case with an algorithm that first splits the intial database into subsets through a weighting procedure). In the case of morphological character states one normally employs non-equivalent (philosophically speaking) types of data simultaneously, such as data pertaining to shape, size, colour and number. Another common assumption of these algorithms based on congruence is the independence of each datum. However, we know that most (if not all) genes influence the development of several character states (pleiotropy) and most (if not all) character states are influenced by several genes (polygenes).

It is generally assumed that the closest approximation to meeting the assumptions of equivalence and independence of the individual data is encountered with DNA and RNA sequence data. However, even with these data we now have increasing evidence of varying degrees of linkage (e.g. the highly conserved sections of some chromosomes). Furthermore it is increasingly recognised that the problems posed by the functional independence of morphological traits and by developmental integration giving rise to convergent resemblance are not restricted to morphological data. There is a growing list of examples of the same problems with some molecular data (e.g. Lee 1999). Nevertheless, congruence algorithms used to analyze such sequence data are capable of yielding useful clusters based on genetic affinity. As such affinity tends to be closely correlated with phylogenetic affinity (in terms of recency of common origin) such clusters are more often than not likely to be monophyla (monophyletic taxa) in the strictest Hennigian sense. However, there may be cases where a species exhibits a closer genetic affinity with a more distant relative than with its immediate sister species, when the latter has acquired novel genes by horizontal gene transfer (e.g. Avise 1994). With advancing knowledge we should be able to detect such anomalies and exclude them from our analyses.

When it comes to the consideration of cladograms generated by algorithms (such as those included in PAUP) applied to morphological data, the blatant invalidation of the above assumptions becomes more serious. The problems posed by convergent evolution further compound the problem. Härlin apparently assumed that convergent evolution is minimal. He therefore dismissed the evidence to the contrary as reviewed by Moore & Willmer (1997), by misrepresenting their position (that whichever method is employed one is confronted with a result that suggests that convergence must be extremely common) and without any valid argument in support of his dismissal of their evidence for the prevalence of convergent evolution. Within Diptera, for example, independent, convergent losses of character states have evidently been especially commonplace. Furthermore the ontogenetic integration of an organism can mean that a simple morphological change may trigger a cascade of compensatory changes during development. Thus supposedly complex morphological characters may have a relatively simple developmental basis, and therefore be more likely to evolve convergently than one intuitively might suspect (Lee 1999). In short, the amount of misleading noise is highly likely to swamp the phylogenetic signal we are seeking. The result of employing the criterion of congruence (parsimony) is to risk postulating synapomorphies on the basis of circular argument masquerading as objectivity. Nevertheless the resulting cladograms, while not being definitive conclusions, they are still useful as research programmes.

The classic Hennigian method is to select relatively few postulated synapomorphies whose homologies, and whose polarities of the postulated transformation series, are based on explicit evidence and well grounded arguments. The latter will frequently involve plausible evolutionary scenarios and some sort of, albeit informal, ranking of the probabilities of different character states being likely to have evolved more than once. The resulting cladograms can then be readily tested by exposure to new sources of data (molecular, further morphological, behavioural, physiological, etc.).

All cladograms are research programmes not definitive conclusions. They cannot be proved but they can be demolished by fresh data. A cladogram based on the preferential selection of character states that are the most plausible postulated synapomorphies (in relation to the available evidence and best argued inferences) is one whose basis is not only fully explicit but it can be readily critically evaluated point by point (bifurcation by bifurcation or node by node) independently. It can thus be progressively revised and refined in the light of revised interpretations and new data. All the time, however, it remains both a research programme and a concise summary of our current interpretations of the most thoroughly evaluated data pertaining to the taxa in question.

By contrast a cladogram based on many data and congruence, as the criterion for the rejection of many data in our initial data base, is seemingly less useful in directing attention to the significance of new data. This is because the inclusion of any datum in the final cladogram is based solely on its congruence with the largest subset of character states in the initial data base. It therefore follows that when one wishes to incorporate new data into a fresh analysis it is imperative to include all the data rejected by one's previous analysis if one is to avoid progressive magnifications of errors in the original result. The fresh analysis (based on congruence) is likely to favour some data previously rejected and to reject some previously favoured. That is the nature of the operation of the congruence criterion.

## Conclusion

I repeat that in practice character states are the starting point for producing a cladogram, which is a summary of interlinked hypotheses as to the form of the evolutionary tree that we aspire to discover. All cladograms are based on selected character states. The criterion employed for this preferential selection must be explicit. It is now generally agreed that this criterion is synapomorphy. I reiterate that the currently debated issue is – how do we recognize synapomorphies?

An attraction of the use of congruence to identify synapomorphies retrospectively is the fact that it involves minimal understanding of the character states selected, it involves no hypotheses as to their significance and it utilises many data. Another advantage is that modern computers can handle a very large number of character states and taxa very rapidly to produce a cladogram (or a set of cladograms). Its disadvantage is that it is difficult to evaluate apart from adding yet more data to the original data base and repeating the exercise to see how consistent the new cladogram is with the previous one. However, where these cladograms differ we frequently have little idea as to which version is nearer the truth. The addition of new data may have merely further obscured the true signal with additional noise. By contrast the classic Hennigian approach, by progressively modifying or replacing cladograms based on many individual explicit arguments, is perhaps more likely to be progressively approximating the true evolutionary tree. Furthermore each datum on which it is based involves explicit hypotheses advanced prior to the generation of the cladogram. These hypotheses are available for testing at any time one by one. An apparent disadvantage of this method is the smaller number of data on which the cladogram is based. However, it is the common experience to find that many character states not employed in the construction of the cladogram are subsequently found to be congruent with it. Thus the classic Hennigian approach can reveal the pattern of evolution of these, often seemingly trivial, character states as well as revealing convergent evolution among many others. The classic Hennigian approach, therefore, would still seem to be useful as a research programme.

In practice it is normally found that there is a, perhaps surprising, measure of agreement between cladograms generated by the two approaches. Where they differ, therefore, can usefully highlight areas that merit further research – rather than dogmatic declarations regarding the superiority of one's chosen approach.

I conclude that the highly partisan advocacy of only one approach to producing a cladogram, such as Härlin's advocacy of the use of congruence for the retrospective recognition of synapomorphies, contributes nothing new and may mislead by offering an apparently novel, but essentially spurious, defence of this one chosen method. Every method rests on dubious assumptions and/or limited data. As with imperfect population estimates (such as the Lincoln Index) in ecology proving very useful for comparisons between different populations in space and time, so imperfect procedures for generating cladograms are useful procedures for directing future research. Such research may either disprove all or part of our favoured cladogram. It may fail to disprove it. We live with those cladograms we fail to disprove. There is no panacea procedure for discovering the true cladogram.

## Acknowledgements

I am currently funded by the Isaac Newton Trust (Trinity College, Cambridge). I am grateful to my colleague Dr Janet Moore for kindly commenting on an initial draft of this note and to the referees for helpful comments.

## Zusammenfassung

Ist die Hennigsche Methode zur Erstellung von Kladogrammen nicht ebenso vertretbar wie diejenigen, die auf Parsimonie-Algorithmen basieren?

In einem kürzlich veröffentlichten Beitrag argumentierte Härlin (1999), dass in der phylogenetischen Systematik der für ein Taxon postulierte Stammbaum logisch Priorität über taxonomische Merkmale hat und dass sich daraus ableitet, dass die Kongruenz von Merkmalen und Merkmalszuständen (mit diesem Stammbaum) Prioriät über alle anderen Überlegungen hat. Diese Auffassung wird hier kritisch diskutiert. Die Methode von Hennig (1966) zur Begründung von Stammbäumen wird als zumindest ebenso nützlich zur Formulierung von sinnvollen und überprüfbaren Hypothesen angesehen.

## References

Avise, J. C. (1994): Molecular Markers, Natural History and Evolution. – Chapman & Hall, London, 511 pp.

Disney, R. H. L. (1983): A synopsis of the taxonomist's tasks, with particular attention to phylogenetic cladism. – Field Studies 5(5): 841–865.

Disney, R. H. L. (1999): Insect biodiversity and the demise of alpha taxonomy. – Antenna 23: 84–88.

Härlin, M. (1999): The logical priority of the tree over characters and some consequences for taxonomy. – Biol. J. Linn. Soc. 68: 497–503.

Hennig, W. (1966): Phylogenetic Systematics. – University of Illinois Press, Urbana, 263 pp.

Lee, M. S. Y. (1999): Molecular phylogenies become functional. – Trends Ecol. Evol. 14: 177–178.

Mayr, E. & Ashlock, P. D. (1991): Principles of Systematic Zoology. – McGraw-Hill, New York, 475 pp.

Moore, J. & Willmer, P. (1997): Convergent evolution in invertebrates. – Biol. Rev. 72: 1–60.

Scotland, R. W. (1992): Cladistic theory. – In: Forey, P. L., Humphries, C. J., Kitching, I. J., Scotland, R. W., Siebert, D. J. & Williams, D. M., editors. Cladistics. Clarendon Press, Oxford. Chapter 1: 3–13.

Wägele, J.-W. (1996): First principles of phylogenetic systematics, a basis for numerical methods used for morphological and molecular characters. – Vie Milieu 46: 125–138.

Yeates, D. K. & Wiegmann, B. M. (1999): Congruence and controversy: toward a higher-level phylogeny of Diptera. – Ann. Rev. Ent. 44: 397–428.

Dr R. H. L. Disney, Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, England.